9К

УДК 517.958:531.33

Ю.В. Крак, І.О. Стеля
Київський національний університет імені Тараса Шевченка, м. Київ, Україна
krak@unicyb.kiev.ua, Igor.Stelia@gmail.com

Комп'ютерна модель голосових зв'язок та мовного тракту людини

Розроблені математичне та програмне забезпечення для моделювання та аналізу голосових зв'язок та мовного тракту людини, що разом складають підґрунтя для систем синтезу мови на основі математичних моделей.

Вступ

Метою статті є розробка математичного і програмного забезпечення для моделювання і аналізу голосових зв'язок та мовного тракту людини. Голосові зв'язки є джерелом звуків, які створюються людиною, і спільно з моделями мовного тракту складають підґрунтя для систем синтезу мови на основі математичних моделей. Таким чином, створення комп'ютерних моделей є актуальним для вивчення мовотворення.

Математичні моделі

За основу математичної моделі голосових зв'язок обрана двомасова модель [1], згідно з якою кожна з голосових зв'язок описується двома масами, що здійснюють автоколивання під впливом потоку повітря в голосовій щілині. Система рівнянь для двох мас, що коливаються, записується у вигляді:

$$\begin{aligned} m_1 \frac{d^2 x_1(t)}{dt^2} + r_1 \frac{dx_1(t)}{dt} + k_1(x_1(t) - x_{01}) + k_c(x_1(t) - x_2(t)) = l_g d_1 p_{m1}(t), \\ m_2 \frac{d^2 x_2(t)}{dt^2} + r_2 \frac{dx_2(t)}{dt} + k_2(x_2(t) - x_{02}) - k_c(x_1(t) - x_2(t)) = l_g d_2 p_{m2}(t), \end{aligned} \tag{1}$$

де m_1 і m_2 – маси, $x_1(t)$, $x_2(t)$ – зміщення мас m_1 і m_2, x_{01}, x_{02} – початкові положення мас m_1 і m_2, t – час, r_1 і r_2 – коефіцієнти демпфування, k_1, k_2 – пружність пружин для мас m_1 і m_2, k_c – пружність пружини, що з'єднує маси m_1 і m_2, d_1 і d_2 – товщина мас m_1 і m_2, l_g – діюча довжина голосових зв'язок, $l_g d_1$, $l_g d_2$ – поверхні мас m_1 і m_2, на які діють тиски $p_{m1}(t)$ і $p_{m2}(t)$.

Система (1) розв'язується спільно з системою рівнянь для змін тиску (будується на основі рівнянь Бернуллі), які відбуваються в потоці при його проходженні через голосові зв'язки.

В результаті реалізації моделі визначається потік повітря $u_g(t)$ на виході з голосової щілини. Цей потік використовується як вихідні дані для моделі мовного тракту.

Для моделювання розповсюдження акустичних хвиль у мовному тракті як у неоднорідній акустичній трубі, що починається між голосовими зв'язками та закінчується губами, використовується система лінійних рівнянь в частинних похідних [2], [3], яка записується у вигляді:

$$-\frac{\partial p}{\partial x}=\frac{\rho}{A(x)}\frac{\partial u}{\partial t},$$
$$-\frac{\partial u}{\partial x}=\frac{A(x)}{\rho c^2}\frac{\partial p}{\partial t}, \quad (2)$$

де $p(x,t)$ – тиск у тракті в точці x ($0 \le x \le L$, де L – довжина мовного тракту) в момент часу t, $u(x,t)$ – об'ємна швидкість потоку, ρ – густина повітря в тракті, c – швидкість звуку і $A(x)$ – функція площі поперечного перетину.

Оскільки тракт має неоднорідний поперечний перетин, він розбивається на циліндричні секції однакової довжини з постійною площею перетину.

В якості крайової умови на вході в тракт вибирається потік $u_g(t)$, знайдений в моделі голосових зв'язок. Виходячи з цього, маєму крайову умову:

$$u(0,t)=u_g(t).$$

На протилежному кінці тракту задаємо умову

$$p(L,t)=0.$$

Програмне забезпечення

В основу алгоритму для реалізації моделі голосових зв'язок [4] покладено метод Кутта-Мерсона, який об'єднується в єдину ітераційну процедуру з методом послідовної верхньої релаксації для розв'язування нелінійною системою для зміни тиску. Для розв'язування системи (2) використовується метод «чехарди» [5], [6].

Складність фізичного явища, що вивчається, призводить до необхідності урахування великої кількості параметрів. В свою чергу, математична модель є значною мірою чутливою до збурень цих параметрів. Для зручного введення та редагування вихідних параметрів моделі та відображення результатів моделювання був розроблений графічний інтерфейс користувача. Екранна форма для введення параметрів моделі голосових зв'язок наведена на рис. 1. За допомогою програмного забезпечення користувач моделі може вивчати вплив параметрів моделі на параметри голосового джерела, не будучи фахівцем з математичного моделювання. Необхідні для вивчення джерела параметри і результати моделювання відображаються у цифровому та графічному вигляді за допомогою екранних форм користувача.

Для вивчення голосового джерела, окрім функції потоку, використовується його похідна. Приклад відображення результатів моделювання потоку та його похідної наведений на рис. 2.

Набір параметрів для кількісного опису коливань вводиться із залежності похідної потоку від часу і включає [7]: амплітуду від'ємного піка, тривалість фази відкриття, тривалість закритої фази, проміжок часу між додатним і від'ємним піками, час повернення в початкове положення і вторинні параметри, що походять з них:

швидкість відкриття, тривалість фази відкриття голосової щілини відносно періоду основного тону, тривалість фази відкритої голосової щілини відносно тривалості основного тону тощо. Ступінь змикання голосових зв'язок (відстань між зв'язками) відіграє важливу роль у визначенні тривалості фази закриття голосової щілині, різкості закриття і впливає на якість сприйняття фонації.

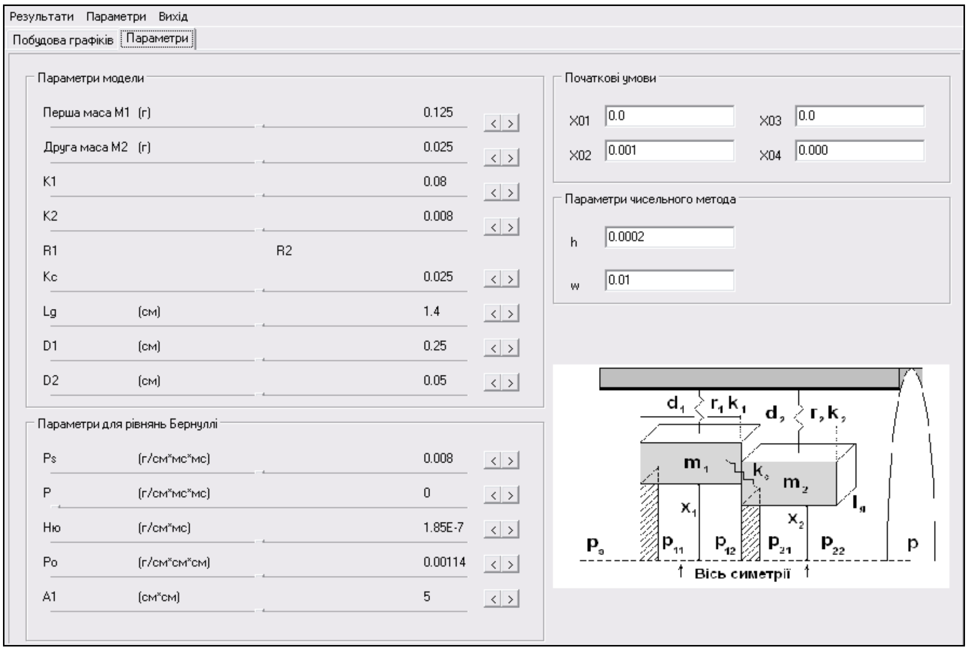


Рисунок 1 – Екранна форма для введення параметрів моделі голосових зв'язок

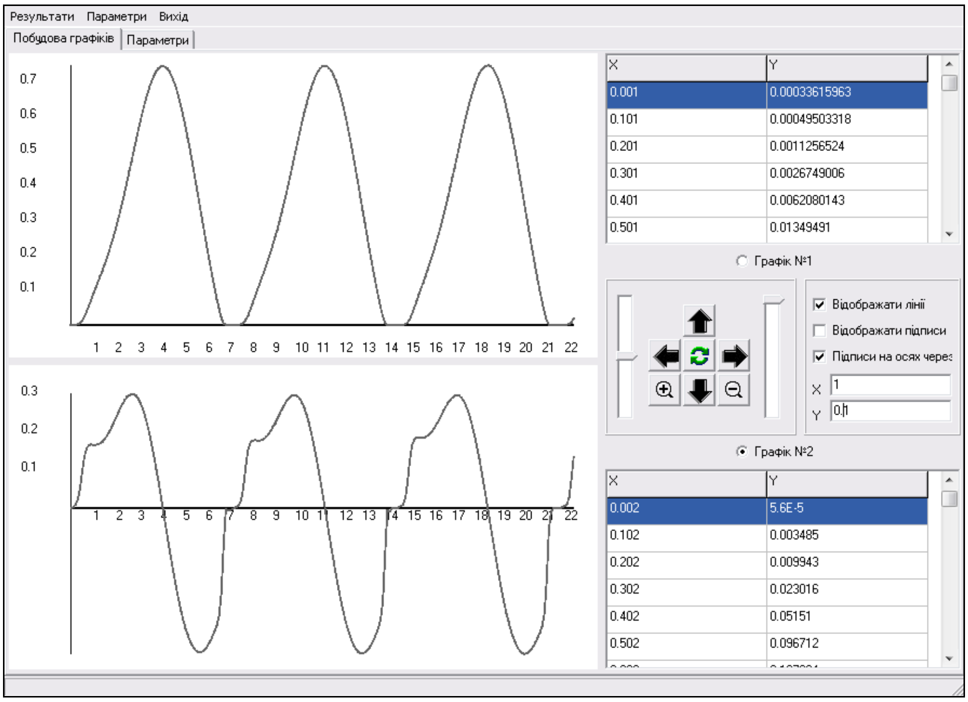


Рисунок 2 – Результати моделювання голосового джерела у графічному вигляді

Набір параметрів для кількісного опису коливань вводиться із залежності похідної потоку від часу і включає [7]: амплітуду від'ємного піка, тривалість фази відкриття, тривалість закритої фази, проміжок часу між додатним і від'ємним піками, час повернення в початкове положення і вторинні параметри, що походять з них: швидкість відкриття, тривалість фази відкриття голосової щілини відносно періоду основного тону, тривалість фази відкритої голосової щілини відносно тривалості основного тону тощо. Ступінь змикання голосових зв'язок (відстань між зв'язками) відіграє важливу роль у визначенні тривалості фази закриття голосової щілині, різкості закриття і впливає на якість сприйняття фонації.

Для введення і редагування вихідних даних моделі мовного тракту, зокрема, задання геометричної форми тракту та відображення результатів моделювання вихідного мовного сигналу у цифровому та графічному вигляді був розроблений графічний інтерфейс користувача, екранна форма якого наведена на рис. 3. За його допомогою користувач моделі може задавати бажану форму мовного тракту та вивчати її вплив на вихідний мовний сигнал. На рис. 3 наведені форма тракту (верхній графік) та вихідний сигнал (нижній графік) для російського голосного «а». Параметри мовного тракту для російських голосних були використані з роботи [8].

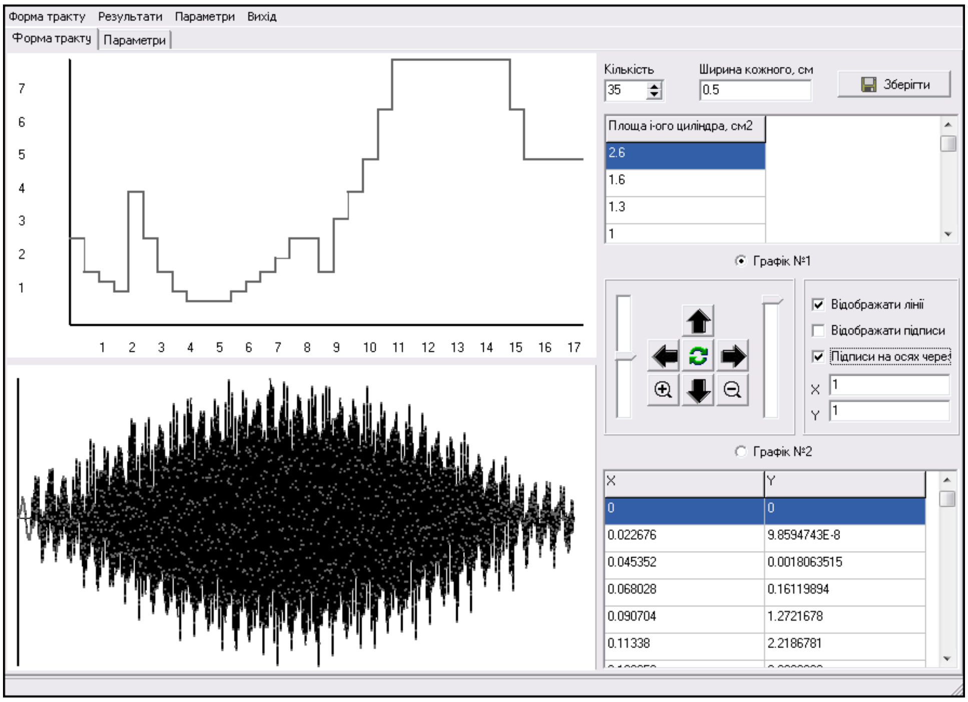


Рисунок 3 – Екранна форма для введення параметрів тракту
і візуалізації вихідного сигналу

Висновки

Розроблені в роботі чисельні алгоритми та програмне забезпечення дають змогу синтезувати і досліджувати мовний сигнал, використовуючи на вході в мовний тракт потік, розрахований за моделлю голосових зв'язок. Розроблений інтерфейс користувача дозволяє у зручному режимі вивчати вплив параметрів голосових зв'язок та

форми мовного тракту на вихідний сигнал. Наведені в роботі результати моделювання добре узгоджуються з існуючими на цей час уявленнями про джерело звуків, яким є голосові зв'язки людини, та з наведеними в літературі результатами [9-11], одержаними за іншими моделями голосових зв'язок. Якість сигналу на виході з мовного тракту оцінюється експертами безпосередньо при відтворенні відповідних звуків.

Література

1. Ishizaka K. and Flanagan J.L. Synthesis of voiced sounds from a two-mass model of vocal cords // Bell Syst. Tech. J. – 1972. – 51(6). – P. 1233-1268.
2. Portnoff M.R. A Quasi-One-Dimensional Digital Simulation for the Time-Varying Vocal-Tract // S.B./S.M. thesis, MIT. – Cambridge, Mass. – 1973.
3. J. Schroeterand M.M. Sondhi. Techniques for Estimating Vocal-Tract Shapes from theSpeech Signal // IEEETrans. Speech Audio Proc. – 1994. – Vol. 2(1). – P. 133-150.
4. Крак Ю.В., Стеля І.О. Чисельне моделювання голосових зв'язок за двомасовою моделлю // Журнал обчислювальної та прикладної математики. – 2007. – № 1(94). – С. 55-60.
5. Randal J. Leveque. Finite volume methods for hyperbolic problems. – Cambridge University Press, 2002.
6. Morton K.W., Mayers D.F. Numerical solutions of partial differential equations. – Cambridge University Press, 2005. – 2nd ed.
7. Alku P. and Vilkman E. A comparison of glottal voice sorce quantification parameters in breathy, normal, and pressed phonation of female and male speakers // Folia Phoniatr. Logop. – 1996. – № 48. – P. 240-254.
8. Fant C.G.M. Acoustic Theory of Speech Production. – The Hague: Mouton and Co, 1960.
9. Flanagan J.L. and Landgraf L.L. Self-oscillating source for vocal tract synthesizers // IEEE Trans. Audio Electroacoust. AU-16. – 1968. – P. 57-64.
10. Meyer P., Wilhelms R. and H.W.Strube. A quasiarticulatory speech synthesizer for German language running in real time // J. Acoust. Soc. Am. – 1989. – August. – № 86 (2). – P. 523-539.
11. Childers D., Hicks D., Moor G. and Alsaka Y. A model for vocal fold vibratory motion, contact area, and the electroglottogram // J. Acoust. Soc. Am. – 1986. – № 80. – P. 1309-1320.

Ю.В. Крак, И.О. Стеля

Компьютерная модель голосовых связок и речевого тракта человека

Разработаны математическое и программное обеспечение для моделирования и анализа голосовых связок и речевого тракта человека, которые вместе составляют основу для систем синтеза речи, построенных на математических моделях.

Yu.V. Krak, I.O. Stelya

Computer Models of Human Vocal Folds and Vocal Tract

Mathematical facilities and software for human vocal folds and vocal tract models are developed. The models constitute a basis for speech synthesis systems.

Стаття надійшла до редакції 24.07.2008.